



Guía docente				
Datos Identificativos				2022/23
Asignatura (*)	Probabilidad. estadística y elementos de biomatemática	Código	614522007	
Titulación	Mestrado Universitario en Bioinformática para Ciencias da Saúde			
Descriptorios				
Ciclo	Periodo	Curso	Tipo	Créditos
Máster Oficial	1º cuatrimestre	Primero	Obligatoria	6
Idioma	CastellanoGallegoInglés			
Modalidad docente	Presencial			
Prerrequisitos				
Departamento	Matemáticas			
Coordinador/a	Cao Abad, Ricardo	Correo electrónico	ricardo.cao@udc.es	
Profesorado	Cao Abad, Ricardo	Correo electrónico	ricardo.cao@udc.es	
Web	http://dm.udc.es/staff/ricardo_cao/			
Descripción general	Se pretende que el alumno adquiera destreza en la identificación de situaciones en las que el la teoría de la probabilidad y los métodos de inferencia estadística son herramientas adecuadas para el análisis cuantitativo de bases de datos generadas en el ámbito de la Bioinformática. Para ello se tratará de que el alumno complemente su conocimiento sobre los conceptos básicos de probabilidad e inferencia estadística, obtengan soltura en el manejo del software estadístico R, utilizando un importante número de recursos e introduciendo al estudiante en la programación en este entorno. Asimismo se persigue que el alumno se familiarice con los modelos probabilísticos de procesos estocásticos en tiempo discreto y adquiera una formación básica en técnicas de remuestreo (bootstrap) como herramienta para la aplicación y evaluación de diferentes algoritmos estadísticos.			

Competencias / Resultados del título	
Código	Competencias / Resultados del título
A5	CE5 - Desarrollo de habilidades en el manejo de técnicas estadísticas y su aplicación a conjuntos de datos del campo de la Bioinformática
A6	CE6 - Capacidad para identificar las herramientas software y fuentes de datos de bioinformática más relevantes, y adquirir destreza en su uso
A10	CE10 - Elaborar un proyecto de investigación bioinformática, anticipando obstáculos y las posibles estrategias alternativas para solucionarlos.
B1	CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
B4	CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades
B5	CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida auto dirigido o autónomo.
C3	CT3 - Utilizar las herramientas básicas de las tecnologías de la información y las comunicaciones (TIC) necesarias para el ejercicio de su profesión y para el aprendizaje a lo largo de su vida
C6	CT6 - Valorar críticamente el conocimiento, la tecnología y la información disponible para resolver los problemas con los que deben enfrentarse
C8	CT8 - Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad

Resultados de aprendizaje			
Resultados de aprendizaje		Competencias / Resultados del título	
G2 - Capacidad de aplicación de algoritmos de resolución de los problemas y manejo del software adecuado.		AP5	BP1
		AP6	CP3
		AP10	



G1 - Capacidad para iniciar la investigación y para participar en proyectos de investigación que pueden culminar en la elaboración de una tesis doctoral.	AP5 AP6 AP10	BP1 BP4 BP5	CP3 CP6 CP8
G3 - Capacidad de trabajo en equipo y de forma autónoma	AP5 AP6	BP1 BP4 BP5	CP3 CP6 CP8
G4 - Capacidad de formular problemas en términos estadísticos, y de resolverlos utilizando las técnicas adecuadas.	AP5 AP6 AP10	BP1	CP3 CP6
G6 - Capacidad de identificar y resolver problemas	AP5 AP6 AP10	BP1 BP5	CP3
G10 - Capacidad de integrarse en un equipo multidisciplinar para el análisis experimental	AP5 AP6 AP10	BP1 BP4 BP5	CP3 CP6 CP8
G11 - Adquirir destreza para el desarrollo de software	AP5 AP6	BP5	CP3
G12 - Capacidad de análisis estadístico crítico de las muestras, los planteamientos y resultados	AP5 AP10	BP1 BP5	CP6 CP8
G14 - Representar un problema real mediante un modelizado estadístico adecuado.	AP5 AP6 AP10	BP1 BP5	
G15 - Diseñar un plan de observación o recogida de datos que permita abordar el problema de interés	AP5 AP6 AP10	BP1 BP5	CP3 CP6
E2 - La adquisición de los conocimientos de estadística e investigación operativa necesarios para la incorporación en equipos multidisciplinares pertenecientes a diferentes sectores profesionales.	AP5 AP6 AP10	BP1 BP4 BP5	CP3 CP6 CP8
E4 - Conocer las aplicaciones de los modelos de la estadística y la investigación operativa.	AP5 AP10	BP1 BP4 BP5	CP6
E5 - Conocer algoritmos de resolución de los problemas y manejar el software adecuado.	AP5 AP6 AP10	BP1 BP5	CP3 CP6 CP8
E12 - Realizar inferencias respecto a los parámetros que aparecen en el modelo.	AP5 AP6 AP10	BP1 BP4 BP5	CP3 CP6 CP8
E19 - Tratamiento de datos y análisis estadístico de los resultados obtenidos.	AP5 AP6 AP10	BP1 BP4 BP5	CP3
E27 - Obtener los conocimientos precisos para un análisis crítico y riguroso de los resultados.	AP5 AP10	BP1 BP4 BP5	CP6 CP8
E28 - Complementar el aprendizaje de los aspectos metodológicos con apoyo de software.	AP6 AP10	BP5	CP3 CP6 CP8
E78 - Fomentar la sensibilidad hacia los principios del pensamiento científico, favoreciendo las actitudes asociadas al desarrollo de los métodos matemáticos, como: el cuestionamiento de las ideas intuitivas, el análisis crítico de las afirmaciones, la capacidad de análisis y síntesis o la toma de decisiones racionales	AP5 AP10	BP1 BP4 BP5	CP6 CP8



E82 - El estudiante será capaz de comprender la importancia de la Inferencia Estadística como herramienta de obtención de información sobre la población en estudio, a partir del conjunto de datos observados de una muestra representativa de ésta. Para ello deberá reconocer la diferencia entre estadística paramétrica y no paramétrica.	AP5 AP10	BP1 BP4 BP5	CP6 CP8
E84 - Ser capaz de manejar diverso software (en particular R) e interpretar los resultados que proporcionan éstos en los correspondientes estudios prácticos.	AP5 AP6 AP10	BP4 BP5	CP3
E86 - Soltura en el manejo de la teoría de la probabilidad y las variables aleatorias.	AP5 AP10	BP1 BP4 BP5	CP6

Contenidos	
Tema	Subtema
1. Revisión de conceptos básicos de probabilidad y estadística.	a. Probabilidad. Variables aleatorias y distribuciones notables discretas y continuas. Distribuciones multivariantes. b. Inferencia estadística: estimación, contrastes de hipótesis e intervalos de confianza.
2. Revisión del lenguaje de programación estadístico R.	a. Introducción al R. Primeros pasos. Funciones internas. Ayuda en R. Funciones, bucles, vectores. Funciones estadísticas. Gráficas. Recursividad. R studio. b. Principales distribuciones de probabilidad en R. c. Introducción a la simulación en R. d. Estadística descriptiva en R. e. Contrastes de hipótesis e intervalos de confianza con R.
3. Modelos estadísticos lineales.	a. El modelo de regresión lineal simple. Hipótesis básicas. Estimación. Contrastos. Predicción. Diagnóstico del modelo. b. El modelo de regresión lineal múltiple. Hipótesis básicas. Estimación. Contrastos. Predicción. Diagnóstico del modelo. c. Modelos básicos del diseño experimental. Análisis de la Varianza (ANOVA) de una y dos vías, sin y con interacción. Hipótesis básicas. Estimación. Contrastos. Diagnóstico del modelo. d. El problema de los contrastes múltiples. False discovery rate.
4. Introducción a los procesos estocásticos.	a. Paseo aleatorio simple. b. Proceso de Poisson y procesos de renovación. Procesos de nacimiento y muerte. c. Procesos Markovianos. Cadenas de Markov.
5. Introducción a los métodos de remuestreo.	a. El bootstrap uniforme. Cálculo de la distribución bootstrap: distribución exacta y distribución aproximada por Monte Carlo. Ejemplos. Aplicación del bootstrap a la estimación de la precisión y el sesgo de un estimador. b. Modificaciones del Bootstrap uniforme. Bootstrap paramétrico, simetrizado y suavizado. Discusión y ejemplos. c. Métodos bootstrap para la construcción de intervalos de confianza: método percentil, percentil-t, percentil-t simetrizado. Ejemplos. Estudios de simulación.

Planificación				
Metodologías / pruebas	Competencias / Resultados	Horas lectivas (presenciales y virtuales)	Horas trabajo autónomo	Horas totales
Presentación oral	A5 A6 A10 B1 B4 B5 C8	24	36	60



Prácticas a través de TIC	A5 A6 A10 B4 B5 C3 C6	18	36	54
Prueba de respuesta múltiple	A5 B1 B5 C8	1	9	10
Solución de problemas	A5 A6 A10 B1 B4 B5 C3 C6 C8	4	16	20
Atención personalizada		6	0	6

(*) Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos

Metodologías	
Metodologías	Descripción
Presentación oral	Presentación con ordenador.
Prácticas a través de TIC	Análisis estadístico de conjuntos de datos usando R.
Prueba de respuesta múltiple	Prueba de repuesta múltiple sobre conceptos.
Solución de problemas	Elección de las herramientas estadísticas y estrategias para resolver problemas. Formulación de modelos lineales. Diseño de experimentos. Formulación de planes de remuestreo.

Atención personalizada	
Metodologías	Descripción
Prácticas a través de TIC	Asistencia y participación en las clases teóricas. Examen escrito de múltiple opción.
Solución de problemas	Participación en prácticas y seminarios. Supuesto práctico a realizar por el alumno.

Evaluación			
Metodologías	Competencias / Resultados	Descripción	Calificación
Presentación oral	A5 A6 A10 B1 B4 B5 C8	Presentación oral del trabajo original mencionado en el item "Solución de problemas".	10
Prácticas a través de TIC	A5 A6 A10 B4 B5 C3 C6	Práctica de ordenador usando el software estadístico libre R.	30
Solución de problemas	A5 A6 A10 B1 B4 B5 C3 C6 C8	Trabajo original sobre alguno de los temas de la materia en un contexto de interés en Bioinformática.	30
Prueba de respuesta múltiple	A5 B1 B5 C8	Prueba de comprensión de los conceptos impartidos.	30

Observaciones evaluación



La evaluación se realizará por medio de una prueba sobre prácticas con R, un trabajo individual del/a alumno/a, así como una prueba escrita de conceptos. La calificación de la prueba de conceptos representará el 30% de la calificación global, la prueba de prácticas en R corresponderá al 30% mientras que el 40% restante corresponderá al trabajo individual, que ha de ser presentado en público por los alumnos. La cuarta parte de la puntuación de este trabajo individual (10% de la calificación global) corresponde a la presentación oral del mismo.

Para superar la materia será necesario obtener una calificación de, por lo menos, 5 sobre 10 en el conjunto de la materia.

En la oportunidad de julio, los alumnos podrán liberarse de hacer las pruebas correspondientes en las que su calificación en la

oportunidad de enero haya sido de, por lo menos, 4 sobre 10.

En la primera oportunidad (enero-febrero), solo los alumnos que se hayan presentado a ninguna de las pruebas evaluables que figuran arriba obtendrán la calificación de NO PRESENTADO. En julio, obtendrán la calificación de NO PRESENTADO los alumnos que no se hayan presentado al examen final de esa fecha.

Fuentes de información

Básica	<ul style="list-style-type: none"> - Cao Abad, R., Francisco Fernández, M., Naya Fernández, S., Presedo Quindimil, M.A., Vázquez Brage, M (2001). Introducción a la Estadística y sus Aplicaciones. Pirámide - Ewens, W.J. and Grant, G.R. (2005). Statistical Methods in Bioinformatics. Springer - Peña Sánchez de Rivera, D. (2000). Estadística: Modelos y Métodos. Alianza Editorial - Ross, S.M. (1995). Stochastic Processes. Wiley - Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Chapman and Hall - Davison, A.C. and Hinkley, D.V. (1997). Bootstrap Methods and their Application. Cambridge University Press
Complementaria	

Recomendaciones

Asignaturas que se recomienda haber cursado previamente

Asignaturas que se recomienda cursar simultáneamente

Introducción a las bases de datos/614522002

Genómica/614522006

Fundamentos de bioinformática/614522008

Introducción a la programación/614522001

Fundamentos de inteligencia artificial/614522003

Asignaturas que continúan el temario

Estructuras de datos y algoritmia para secuencias biológicas/614522013

Procesamiento avanzado de secuencias biológicas/614522020

Inteligencia computacional para datos de alta dimensionalidad/614522024

Trabajo fin de máster/614522025

Inteligencia computacional para bioinformática/614522012

Métodos estadísticos avanzados en bioinformática/614522009

Otros comentarios

(*) La Guía Docente es el documento donde se visualiza la propuesta académica de la UDC. Este documento es público y no se puede modificar, salvo cosas excepcionales bajo la revisión del órgano competente de acuerdo a la normativa vigente que establece el proceso de elaboración de guías